Individual Review of *Fisheries Stock Assessments for Arrowtooth Flounder, Flathead Sole and Kamchatka Flounder*

Prepared for:
The Center for Independent Experts

June 2017

Dr Kevin Stokes
Stokes.net.nz Ltd
59 Jubilee Rd
Khandallah
Wellington 6035
New Zealand
Ph: +64 (04) 973 7305
E-mail: kevin@stokes.net.nz

Executive Summary

The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

Assessments for three Bering Sea and Aleutian Islands (BSAI) flatfish stocks (arrowtooth flounder (*Atheresthes stomias*), flathead sole (*Hippoglossoides elassodon*), and Kamchatka flounder (*Atheresthes evermanni*)) were reviewed. All stocks are subject to individual TACs and are caught in directed and other fisheries. Since 2008, all have been primarily caught by the Amendment 80 fleet. Jointly, the three stocks comprise about 12% of the total BSAI flatfish catches and a little over 1% of total BSAI catches.

Materials sent in advance were limited to the latest, relevant SAFE report chapters; these provided good background. Presentations provided while at the review contained materials not sent in advance. For each stock, the time for presentation and discussion was a half day. The AFSC-provided ToR varied by stock and presentations were generally relevant to those ToR but were not structured specifically to address them. The quality of presentations was high and the openness of presenters and other participants was excellent. It is not clear that public notification of the review was made and no public or stakeholders were physically present during the review. One industry stakeholder did participate via teleconference on the first day.

None of the stocks are defined as overfished or experiencing overfishing and all are subject to low exploitation levels, well below F_{ABC} levels. The assessments all utilize excellent survey data but all suffer to some degree from uncertain catch histories, relatively small amounts of age composition data, or other factors. All assessments therefore have one or more issues that analysts are actively seeking to address, with feedback between analysts, the Plan Team, and SSC. While it is not possible for all technical issues to be resolved given data availability, all assessments appear to provide reliable point estimates of biomass, B40%, F40% and F35% as required for management using Tier 3a by the North Pacific Fisheries Management Council. Reliability is judged on a basis of robustness and consistency.

Some specific recommendations are made in the sections for each stock. None are major and none undermine the conclusion that the science reviewed is the best scientific information available to inform management.

Background

The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

General background to the Bering Sea and Aleutian Islands (BSAI) fisheries was presented by Wilderbuer (Ref 7) and to the Amendment 80 fishery by Haynie (Ref 15). These background papers were useful for the review and were supplemented by additional materials in all stock specific presentations as well as those on the three surveys, ageing, and observers (Refs 8-14).

While there are directed flatfish fisheries with targeting for the species under review, those species are caught both as specified target species by trip and within other trips (targeting other species), as well as in other demersal fisheries targeting species such as Pacific Cod. Fisheries are constrained by the overall North pacific cap of 2 million mt, bycatch in non-flatfish fisheries, catch in non-target flatfish fisheries and by prohibited species catch (PSC) apportionment to multiple fisheries.

Flatfish fishing is governed primarily under the Amendment 80 provision adopted by the North Pacific Fishery Management Council in 2006, and implemented in late 2007. The Amendment allocates several BSAI species among trawl fishery sectors. Catch and PSC allocations by species are made to cooperatives (see, e.g., https://alaskafisheries.noaa.gov/sites/default/files/2016a80_interim_alloc.pdf) which then allocate and manage catches amongst their members.

Bering Sea and Aleutian Islands total catches approach the cap of 2 million mt each year but are dominated by pollock (70%) and Pacific Cod (10%). The flatfish species under review are part of a large flatfish complex with catches dominated by yellowfin sole (*Limanda aspera*) and northern rock sole (*Lepidopsetta polyxystra).* Flatfish catches amount to about 10% of the total catch, with arrowtooth flounder (*Atheresthes stomias*) and flathead sole (*Hippoglossoides elassodon*) contributing 5% each of that 10% (i.e., about one half of one per cent of the total catch) and Kamchatka flounder (*Atheresthes evermanni*) about 2% of the 10% (i.e., about one fifth of one per cent of the total catch). Despite being low proportions of the total, all fisheries are economically important in their own right, as well as potentially creating TAC constraints on other target species. Total catches of all review species are generally well below TACs. Nevertheless, catches can be constrained by TAC (e.g. directed Kamchatka flounder fishing was prohibited from late May, 2016; https://alaskafisheries.noaa.gov/node/54694).

TACs are set by the North Pacific Fisheries Management Council (Council) after receiving advice on multiple issues/factors, including scientific input from the Assessment Plan Teams and Council SSC on Overfishing Limits (OFL) and on acceptable Biological catch (ABC) (see, e.g., http://www.fpir.noaa.gov/SFD/SFD_regs_acls.html for the general national system). The way in which OFL and ABC are advised depends in the North Pacific region on a well-used Management Tier system. The highest level (Tier 1) requires reliable point estimates of biomass and Bmsy and a pdf of Fmsy. Tier 2 requires reliable point estimates of biomass and Bmsy but only YPR-related proxies for Fmsy. Tier 3 requires a reliable point estimate of biomass and YPR-related biomass and fishing mortality reference points.

Tier 3 splits in to three alternatives for setting OFL and ABC depending on stock status. Currently, all three species under review are estimated to be neither overfished nor experiencing overfishing, with estimated B/B40% for all being greater than 1. Consequently, all three species are considered under Tier 3a. The OFL are therefore calculated based on estimates of F35%, while ABC are calculated based on estimates of F40%.

The three flatfish species under review are all currently managed under Tier 3a and stock assessments need to be reviewed with the Tier requirements in mind. Always of interest in reviews is what is meant in the Tier definitions by "reliable". The term relates to a point estimate of biomass (and Bmsy at Tier 1, and Bmsy and Fmsy at Tier 2), implying a single, chosen model. But in stock assessment there is no one, single model that can wholly be considered valid. It is therefore difficult to disentangle fully *reliability* (i.e., consistently) from *validity*. In my view, review needs to consider whether the estimates from stock assessment used for advising on OFL and ABC are *robust* to model structure and statistical fitting of each model variant. That is, even though a point estimate is required, would the emanating advice change much or in an important manner if alternative models or fitting options were chosen. If results and implications are robust, then they can be considered reliable for management purposes. If changes in model structure, assumptions, weighting, etc., do make changes of substance in management-related outputs, then *reliability* becomes moot.

Review Process and Activities

Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

There is no standardised NMFS review process common to all regional fisheries management arrangements. The process varies by region. In most regional review systems (e.g., SARC, STAR, SEDAR), the processes are highly formalised and require close adherence to Terms of Reference (ToR), with review products of direct relevance to fisheries management decision making processes. Most reviews conducted through the AFSC are less formal and do not lead directly into formal decision making processes, instead being typically used internally within AFSC for further assessment development. The ToR for this review reflect such an informal use. Any comments and suggestions made here are in this context. (NB CIE ToR are coloured blue, while AFSC ToR are coloured green.) I note that any recommendations are coloured red.

Reviews which feed into management decision making processes require public notification and opportunities for input. This review is not part of such a formal process, but I understand it is normal AFSC practice to still make a public notification. My understanding (and I am not certain) is that this was neglected in this case, but that industry bodies were contacted to provide an opportunity to make input. On the first day of the review, during the afternoon, one industry representative (Todd Loomis) joined the review by teleconference. The review team was made aware that relevant industry representatives could be available on request. Overall, for a review of this type, I am comfortable that opportunities and arrangements for public engagement were reasonable.

More formal NMFS reviews conducted through the CIE, for other fisheries centers, include requirements for CIE reviewers to contribute to Summary Reports which become important process outputs that are used in further management processes. Like many (perhaps nearly all) AFSC reviews, this review did not include a requirement for a Summary Report. This conflicts somewhat with CIE ToR, but is not considered problematic in practical terms. From a reviewer perspective, it leaves "loose ends" from the review meeting and creates a different need for further consideration during report writing, but, again, this is not a problem.

The approach does, however, introduce other differences in the conduct of the review which can be confusing and inefficient. For example, materials provided in advance were not the "whole story" as revealed through presentations. The AFSC objectives for the review only became clear during the review and even during each presentation. As presentations were only provided at the start of each session, this created (at least for me) an inefficiency. If the purpose is exploratory then from my perspective more time would be preferable actually to explore model assumptions as well as more careful consideration of model diagnostics as part of a model building exercise.

The issue was further confused somewhat by the different wording in ToR for Kamchatka flounder compared to the other two stocks. For Kamchatka flounder the first AFSC ToR refers to *evaluat[ing] the modelling approach* while for the other two stocks the equivalent ToRs refer to *evaluat[ing] the model*. Some care is needed in framing ToR, preparing relevant background materials, and making presentation materials available in advance, recognising that presentation materials in particular are likely to be modified until the last moment.

Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

The review was chaired on the first morning by Anne Hollowed and then by Sandra Lowe, both from the AFSC. CIE reviewers were Robin Cook, Sven Kupschus, and Kevin Stokes. Materials sent in advance (see Appendix 1) were read by all CIE reviewers. Most of the first morning consisted of background talks on surveys, observers, age reading, and management. A half day was then provided for presentation and discussion of each of the three flatfish stocks under consideration. The morning of the final day was used as a feedback session on work requested during the review. The CIE reviewers spent some of the final afternoon in discussion and in individual report preparation.

I participated in all activities, but am keenly aware that the presentation materials (see Appendix 1) were provided only at the start of the sessions and contained considerable material beyond that provided in advance. Presentations were of a high standard and analysts were open and helpful. However, greater clarity as to purpose (see also above) and earlier availability of presentation materials would have assisted me as a reviewer to contribute more. My sense is that too much was presented in too short a time for adequate digestion and consideration during the meeting.

Terms of Reference by stock

Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

See stock specific commentaries and ToR below.

Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

The review was not set up to result in final, consistent views or opinions. Rather, for the three stock assessments under consideration, the review comprised of presentations to enable discussion and formulation of views after the meeting; it is unclear to what extent the individual panellist's views will converge or diverge.

Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

The AFSC ToR did not call for a Summary Report. The short duration meeting allowed half a day per stock for three stocks and no attempt was made to reach consensus on stock specific ToR.

Stock Specific Commentaries including response to AFSC-specified ToR

The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

Bering Sea Kamchatka Flounder (BSKF)

A partial assessment of BSAI Kamchatka flounder was carried out in 2015. A full assessment was carried out in 2016, as reported in Wilderbuer *et al.* (Ref 4, Chapter 7 of the NPFMC Bering Sea and Aleutian Islands SAFE Report) and presented by Wilderbuer (Ref 11).

BSKF 1. Evaluate stock assessment approach to model the Kamchatka flounder resource using three spatially distinct trawl surveys to provide reliable estimates of productivity, stock status, and statistical uncertainty for management advice.

The ToR asks if the approach of using three, spatially distinct surveys can provide reliable productivity, status, and uncertainty estimates for management advice. I see no fundamental difficulty with the general modelling approach of using three, spatially distinct surveys, each with distinct attributes as outlined in the presentation by Hoff (Ref 10, slide 5). Indeed, given the paucity of fishery data evident in the table below (from Ref 11, slide 5), and problems with apportioning Kamchatka flounder catches prior to 2011 (or 2007), the availability and use of such extensive survey data is the only way of potentially assessing the stock to meet Tier 3 requirements. It is good that the three surveys cover the fishing and nursery grounds and provide extensive length and age samples; without these data there would be no stock assessment.

The key problem with the available data is the short series of fishery length data and the highly-restricted availability of any age data, with no fishery age data at all. The use of survey length-age-weight relationships is critical – reliability of assessment outputs for management advice depends then on how those survey data are used and what robustness testing has been carried out. The issue of catch is considered under this ToR. The key issues in the assessment (assumptions/estimation of M, and on selectivity) are considered in separate ToR.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
shelf survey biomass and size composition	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
slope survey biomass and size composition												X		X				X		X		X				X
Aleutians biomass and size composition	X			X			X			X		X		X		X				X		X		X		X
fishery catch	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
fishery length data																		X	X	X	X	X	X			
slope survey age composition												X										X				
Aleutians age composition																				X						

Kamchatka flounder was not differentiated from arrowtooth flounder in reported catches until 2011 although observers reported proportions of catch by species from 2007. The catch history for Kamchatka flounder is therefore based on reported catches only from 2011, proportions of reported catches from 2007 to 2010, and an assumed constant proportion (that observed in 2007) for all years up to and including 2006. From 2007 to 2010 the proportion of Kamchatka flounder in the combined species catch increased from 10% to 55%, apparently reflecting a growing market and increased targeting. The weekly catch reports for 2016 shows high catches from late April, through May, and then a sudden drop off with apparent bycatch only during the remainder of the year. The sudden drop presumably being due to the prohibition noted above (https://alaskafisheries.noaa.gov/node/54694).

The resulting catch history is given in the SAFE report as Table 7-1:

Total combined catch (t) of arrowtooth and Kamchatka flounder in the eastern Bering Sea and Aleutian Islands region, 2001-2006. Catches from 2007 to present, when the two species were differentiated in commercial catches, are reported for Kamchatka flounder only in this table.

year	catch	TAC	ABC	OFL
1991	22,052			
1992	10,382			
1993	9,338			
1994	14,366			
1995	9,280			
1996	14,652			
1997	10,054			
1998	15,241			
1999	10,573			
2000	12,929			
2001	13,908			
2002	11,540			
2003	12,834			
2004	17,809			
2005	13,685			
2006	13,309			
2007	1,183			
2008	6,819			
2009	12,802			
2010	21,153			
2011	9,935	17,700	17,700	23,600
2012	9,514	17,700	18,600	24,800
2013	7772	10,000	12,200	16,300
2014	6,220	7,100	7,100	8,270
2015	4,994	6,500	9,000	10,500
2016	4,533	6,500	9,500	11,000

The resulting catch history for the assessment is unusual, with catches of Kamchatka flounder as 10% of the combined figures, ranging from about 1,000 to 2,000 mt, from 1991 to 2006, building rapidly from 2007 to 2010 to a level previously not seen for the entire, combined fishery since 1991, before dropping off again since then. With no indication of market changes, and no TAC constraints at least in 2011, 2012 and possibly 2013, this is presumably due to other constraints and prohibitions.

The general picture from the commercial catches is consistent with a non-target, low F fishery until 2007 and fresh targeting occurring at a time when apparent good recruitment (based on shelf survey size compositions) became available as recruited (to the fishery) fish on the BS slope and AI. The drop off in catches then appears consistent with recruitment dropping off (though possibly coming to full fishery selectivity in coming years).

Though not inconsistent with the signals in the length compositions and fishery changes, nor with biomass signals from all three surveys, I do think a fuller consideration of the implications of the assumption of constant proportionality on 1991 to 2006 catches is warranted to test the reliability of productivity and status estimates, especially given the paucity of data generally. While a simple variation in the proportion might be tested, attempting to use survey data to estimate historical annual proportions would be preferred. The presentation for the review (at slide 3) shows presence by haul data in the BS shelf surveys from 1982 to 2016 for Kamchatka and arrowtooth flounder. Presence is not a good indicator but presumably it would be possible for all three surveys to estimate proportional biomass from 1991 for Kamchatka flounder *cf* combined catches and to use this with total catch data by area to estimate annual proportions. If those proportions are reasonably constant and of the order of 10%, then no more is needed. But if the proportions vary, then I would **recommend** a simple consideration of an alternative catch data set would be worthwhile to test for reliability.

BSKF 2. Evaluate likelihood profile approach to estimate natural mortality rate (and suggest/provide alternatives?)

Note (in passing): The SAFE report does not mention natural mortality (M) under either of the sections on parameter estimation outside or inside the model, only mentioning it under the section on model evaluation (under results) and generally in the section on projections. This lack of consideration may be an oversight.

Given the amount of survey size and age composition data, and the apparent long periods of low fishing mortality prior to the mid-2000s (though this is unclear), in principle, there may be information in the data to help estimate M. In practice, however, given the general low level of data with which to pin down confounding factors such as selectivity, successful M estimation is *a priori* unlikely. Even in much more data rich assessments, estimation of M is usually fraught with problems and it is much more normal, following explorations, to revert to a fixed M or M schedule and to run sensitivity tests.

Likelihood profiling on M seems rational, but I am not entirely clear on what has been done. The SAFE report notes that profiling over M with catchability fixed for only one of the three surveys gave a value of 0.09. Presumably the slope survey q was fixed, but the value is not reported. Nor is the M profile shown or any diagnostics such as the negative log likelihoods for each data source across the M profile. The presentation during review included an M profile with a minimum at 0.08 but with no information on what was fixed across the profile and again without diagnostics to aid in interpretation.

What was provided in the review, was a simple comparison of the assessment with M fixed at 0.08 and at 0.11 (as used previously and adopted for final advice) showing the implications for F40% and F35% (as used to define ABC and OFL). The differences in the Fx% values are marked and would impact on ABC and hence, potentially, TAC decisions.

In my view, profiling on M is only worthwhile if selectivity can be reasonably well defined and if there is confidence in what else in the assessment can be held constant for the profile runs. However, even then, as in all assessments, each run when automated for the profile is not a true or fair comparison as relative weighting is not fully adjusted across the profile to ensure best fits. Apparent differences in likelihood components and totals can thus be somewhat misleading, although are helpful diagnostically.

More important than profiling is to understand the reliability of management advice given uncertainty about M.

The review presentation, though not the SAFE report, shows three standard methods for estimating a fixed M based on life history traits, as well as the value from a likelihood profile. The source of the previously and still used value of 0.11 is not shown. What is also not shown is any schedule (by age) for natural mortality, as considered for example in Brodziak *et al.* (2009) reporting on a NMFS workshop held at the AFSC. The general advice from that workshop is that a constant M approach may be appropriate when only mature fish are of interest in the assessment but that when juvenile fish are also targeted or caught as bycatch then a size-dependent M should be incorporated in to the assessment. Perhaps more importantly, that workshop also concluded that *alternative models for adult natural mortality patterns should be considered in stock assessment applications, where relevant, [in order to account for this uncertainty].*

Based on the presentation, the fixed and estimated selectivities in the final assessment (Ref 4, Fig 7-10 and Ref 11, slide 22) all peak or asymptote well to the left of the fixed, externally determined maturity ogive, with Mat50 at about age 10 (Ref 4, Fig 7-7). In principle, therefore, unless size-dependent natural mortality decreases very quickly on pre-recruits, the assumption of a constant natural mortality or profiled M could lead to major differences in YPR-related metrics such as B40%, F40% and F35% - the metrics of importance in framing management advice and in framing OFL and ABC.

There was no opportunity during the review to consider assessment runs with alternative constant or profiled natural mortality. I would **recommend** that a size-based method such as Lorenzen (1996; and see references in Brodziak *et al.*, 2009) be applied and that assessment reliability be checked by model runs at 0.08, 0.11 and using the size-dependent natural mortality schedule. The important thing is to test reliability of management advice by considering alternatives.

BSKF 3. Evaluate how survey catchability estimates are derived based on assumptions about relative stock distributions.

Catchability (q) is often simply treated as a scaling parameter to fit data. As such, given the only information on M is also in the survey data, q is aliased with M. If all q's were estimated given a fixed M, then a good starting place for fitting may well be the proportions of biomass estimated in each survey. However, assuming well-behaved models and likelihood surfaces, final estimated q's might well be very different. Given the surveys cover different portions of the stock(s) at different life history stages, and all have different gear and operational attributes (see table below copied from Ref 10, slide 5), there is no *a priori* reason to expect relative stock distributions to be reflected directly by the surveys.

Survey	Survey Design	Depths (m)	Vessels	Sampling Density mean (km^2/haul)	Towing Duration (min)	Towing Speed (knots)	Towing Dynamics	Trawl Net	Doors	Door Connection	Footrope
EBS SLOPE	Random stratified	**200-1200**	**1**	200	30	**2.5**	Dynamic mode	Poly Nor' Eastern	**6 x 9 v 2200 lbs**	**4-point**	**mud sweep gear-8" discs**
EBS SHELF	Fixed stations	20-197	2	1300	30	3	Brakes locked	83-112 Eastern	6 x 9 v 1800 lbs	2-point	Fiber core wire wrapped with rubber fire hose
EBS NORTHERN	Fixed stations	20-100	2	1410	30	3	Brakes locked	83-112 Eastern	6 x 9 v 1800 lbs	2-point	Fiber core wire wrapped with rubber fire hose
ALEUTIAN ISLANDS	Random stratified	20-500	2	157	15	3	Dynamic mode	Poly Nor' Eastern	6 x 9 v 1800 lbs	2-point	Bobbins and Roller Gear
GULF OF ALASKA	Random stratified	20-1000	3	560	15	3	Dynamic mode	Poly Nor' Eastern	6 x 9 v 1800 lbs	2-point	Bobbins and Roller Gear

The problem is when the surfaces have multiple local minima or if the relative stock distributions are used to fix one or more q (as is done for Kamchatka flounder), in which case a more detailed consideration is needed.

Each survey catchability is in reality a reflection of availability (areal and vertical) and of vulnerability (proportion of biomass entering the gear that is retained). Elsewhere, a practical approach to estimating q's has involved the formation of priors based on direct evidence and/or consensus views. See, for example, the STAR panel report for sablefish (2007; http://www.pcouncil.org/wp-content/uploads/STARreport_Sablefish.pdf) or the GoA pollock assessment (2012; https://www.afsc.noaa.gov/refm/docs/2012/GOApollock.pdf).

It would be useful for each survey, if not to develop and use priors for three q's, at least to consider carefully for each survey likely lower and upper bounds on q based on survey attributes and different life stages and behaviours of Kamchatka flounder in each area. I would **recommend** this approach to developing expert-informed q ranges, which may then be used to define a number of sensitivity (reliability) runs.

Overall comment on Kamchatka flounder:

The 2015 update assessment suggested B > B40% by a relatively small margin whereas the 2016 assessment suggests B of about 1.2B40%. This is to be expected given recruitment signals and reductions in catch. On the surface, this suggests Kamchatka flounder should still be classified as a Tier 3a stock. I see no problems with the general modelling approach used, but, in my view, in order to be deemed reliable point estimates (of B, B40%, F40%, and F35%), fuller sensitivity tests need to be run to test the robustness of the point estimates. The clearly emerging issues are the catch history (fixed annual proportion of Kamchatka to to arrowtooth flounder *cf* if possible an estimated proportion from surveys); natural mortality (using also a size-based approach); and alternative formulations for fixing one or more catchability estimate. Any sensitivity runs do of course have to be "tuned" and the task is non-trivial. What priority – if any - should be placed on such work depends on factors beyond this review.

Bering Sea Arrowtooth Flounder (BSAF)

A full assessment of BSAI Kamchatka flounder was carried out in 2016, as reported in Spies *et al.* (Ref 1 Chapter 6 of the NPFMC Bering Sea and Aleutian Islands SAFE Report). That assessment and additional exploratory analyses were presented by Spies (Ref 9).

BSAF 1. Evaluation of the ability of the stock assessment model for arrowtooth flounder, combined with the available data, to provide parameter estimates to assess the current status of arrowtooth flounder in the Bering Sea and Aleutian Islands.

See under ToR BSAF 3.

BSAF 2. Evaluation of the strengths and weaknesses in the stock assessment model for arrowtooth flounder.

See under ToR BSAF 3.

BSAF 3. Evaluation of the assumption that male natural mortality is higher than female in arrowtooth flounder.

The assumption that male arrowtooth flounder natural mortality, M, is higher than for females is based on observations that females are almost always found at higher proportions than males during trawl surveys and that age data indicate females outlive males (Wilderbuer and Turnock, 2009; Ref 3). Annual sex ratio estimates for both Bering Sea surveys suggest male proportions generally in the range 0.3 to 0.4 with some hint even of a declining proportion over time. The Aleutian Islands survey displays the same tendency toward female preponderance though the proportions are reversed in one year (1986). Age data presented for the Gulf of Alaska surveys from 1977 to 2013 suggest the oldest fish caught are males (at 34 years compared to 29 for females) but the proportions at age by year (Ref 9, slide 24) clearly indicate a greater chance of encountering females older than males across all years, at least until the late 2000s when the pattern becomes less clear. That the difference is due to natural mortality is supported by the clear difference in growth patterns, with males growing to a smaller size and therefore expected to experience higher predation mortality. Intriguingly, the proportions at age for both males and females appear to show a trend through time to older fish being caught, as well as statistically significant decreasing length at age for 2-4 year males and age 1-3 year females over the past 30-40 years.

Given clear differences in male and female growth patterns, and relationships between size and depth in relationship to temperature and other environmental factors, interpretation of the survey age data needs to be treated with care, especially given the apparent evening out of maximum age in recent years and the clear and major increase in arrowtooth biomass. Nevertheless, the data available for assessment suggest a potentially significant difference in natural mortality (or encounterability) between males and females that needs to be incorporated in to the stock assessment - either directly on M or as viewed through selectivity and/or catchability. Wilderbuer and Turnock applied four methods to estimating M for males and females. Using catch curve analysis, they estimated male M as 0.39 and female M as 0.24. Using the Chapman-Robson method they estimated male M as 0.16 and female M as 0.14. Using Hoenig's method they estimated male M as 0.22 and female M as

0.18. using stock assessment models, using female M fixed at 0.2, they estimated male M as 0.33 for BSAI and 0.35 for GoA. The presentation provided during the review (Ref 9) includes likelihood profiles on male M for fixed female M of 0.17, 0.20, and 0.24. The likelihood profile slide (no. 110) was not discussed during the presentation but taken at face value, the minima suggest corresponding male M values at about 0.23, 0.26, and 0.29 – all 0.50-0.60 higher than the corresponding female values, and much less than the final, assumed male-female difference of 0.15 used previously. It is not clear how the profiling work led to the adoption of fixed values of 0.20 and 0.35.

The 2016 assessment used a fixed female M of 0.2 with an exploration of a range of male M values, with evaluation of models based on fits to the observed sex ratio and substantially to consideration of selectivity estimation. It is unclear in the SAFE document how the final assessment run was chosen, but it appears from the presentation that the final values of M used are again 0.20 for females and male 0.35 for males, consistent with Wilderbuer and Turnock's assessment, but quite different to the Chapman-Robson or Hoenig-based estimates. The SAFE report (Ref 1) explains how exploratory runs were made with female M fixed at 0.20 and how overall fit diagnostics including sex ratio estimation and male M, but also female selectivity (presumably using the likelihood profiles shown in Ref 9, slide 110, being copied from Wilderbuer and Sample, 2002 [Ref 18]) were used to define the final choice on fixed M values. However, it does not appear that full M profiling was carried out – a difficult task given the interaction with selectivity and sex ratio estimation and the need to keep consistency across the profile. If the profiling has not been reconsidered recently (and it is not clear), and given i) changes through time in apparent proportions at age, and ii) far greater age data availability, I would **recommend** a reappraisal of the work – it is not clear from the evidence that a difference of 0.15 between male and female M is appropriate and its inclusion in the model could seriously impact on selectivity estimation and hence on YPR-related reference points.

From an assessment perspective, of course, M is confounded with selectivity and catchability and estimation is problematic. Judicious exploration of externally provided estimates of M is reasonable, but it is confusing to work with estimates derived from an earlier assessment model. There are clear differences in male and female growth patterns, with females growing considerably larger and the fitted von Bertalanffy growth curves showing substantial departure between males and females from age 5 onwards. Further, the observed variability (from surveys) in length at age means that there is considerable spread in the derived age conversion matrices (previous or new), making estimation of M and selectivity difficult.

The ToR asks for an evaluation of the assumption that male natural mortality is higher than female natural mortality in arrowtooth flounder. The evaluation is straightforward – all age data (proportions and longevity through time) suggest male M is higher than female M. This needs to be caveated by noting there are growth differences and potential spatial separation by sex and that survey sampling may be biased with male encounterability lower than for females. For assessment purposes, however, the data must be fitted through assumptions about M and/or in selectivity fitting. It is not possible to estimate both M and selectivity and fixing M is sensible, as is then using fits to the key data sources (biomass estimates), length and age frequencies, sex ratio, and scrutinising selectivity to determine a final model choice. This is what I understand to be the crux of the presentation during review; the focus was on selectivity but ultimately the purpose is ensure rational M choice.

The ToR asks about an evaluation of the assumption about differential male and female natural mortality, but it is natural to expand consideration as to whether the base case 2016 model or the further analyses presented during review (using widened selectivity bounds and an updated age conversion matrix), are the best choices as a basis for informing management. I think it is clearly appropriate to use the revised age conversion matrix. Before getting too involved in considerations of selectivity fits as discussed during review, it is worth pointing out immediately that both total and female spawning biomass (FSB) and FSB/B0 ratios, recruitment, and fishing mortality (Ref 9, slides 74-76) are only barely impacted by the choice of models presented (though see below on YPR). Further, considering the likelihood components for the different model runs (Ref 9, slide 77), there is little to aid the decision. The 2016 model has a lower overall log likelihood and the differences in models are due to changes in fit in just a few places (I think the weightings are the same and likelihoods can be compared). Interestingly, the modified assessments fit the shelf and slope survey biomasses slightly worse (not a good feature) and selectivity is also a worse fit, apparently because of larger discrepancies in fitting the slope and AI survey age data. In other words, there may be a conflict in using visual inspection of fitted selectivity ogives with *post hoc* justification and explanation as opposed to considering the actual fits. On fits alone, the 2016 assessment, likely with the modified age conversion matrix, would be a clear choice.

In terms of the estimated selectivity ogives (Ref 9, slide 71) changes were introduced because of problems in hitting bounds but also, apparently (though not clearly), due to dissatisfaction with some of the resulting ogives in the 2016 assessment. There is sense in widening the bounds to free up selectivity fitting and the resulting ogives for the surveys certainly appear “well-behaved”, and no longer hit bounds. However, given the differences in male and female growth rates, it is surprising that the resulting selectivity ogives for males and females are effectively the same in two of the surveys. It is not easy to intuit, but it is surprising that male-female differences in distribution and behaviour should so neatly cancel out the marked differences in observed sex ratio and observed proportions and size at age. While noting the lower bound issue for the 2016 assessment model, the resulting survey selectivity ogives appear more as might be expected *a priori* given differences in growth. Perhaps more work is needed on looking at M – is it really consistently so different between male and female or is there also a trend in the difference (declining)? Any trends in M (or encounterability) will confound selectivity estimation which is the only place the model can try to deal with it. I note that with fishery length composition data available it might be possible to consider selectivity blocking. However, I am reluctant to recommend this given the overall low F and stock status; is such complication warranted? Sometimes it is easier just to make a few runs and see how it works, but a first step would be to look at the detailed likelihood components to look for patterns fishery selectivity in recent years.

One thing not considered during review is how model choice might impact on YPR-related reference points and hence on advice. The SAFE report (Ref 1) says that maturity at age follows Stark (2011, Ref 19). Table 6.7 of Ref 1 and slide 148 of the presentation (Ref 9) show that female maturity starts to rise around age 6/7 and reaches 50% between ages 7 and 8, asymptoting at age 10 onwards. With fishery selectivity being fit non-parametrically, it is quite unclear how B40% and Fx% values might be affected by the choice of model. The 2016 model fishery selectivity levels off at about 0.75 by age 5/6 while the assessment including widened bounds has fishery selectivity estimated as a sawtooth, peaking first at age 7/8 and then maximising at 1 at much higher ages.

Overall, the assessment is underpinned by extensive information on the regional ecosystem and how environmental factors influence spatial distribution. Three long-standing surveys provide good biomass indices which can drive the model fits and there is extensive length composition data for both surveys and the fishery. Age composition data are lacking for the fishery but are good for the Bering Sea shelf survey. Some age composition data are available for recent years for the slope and AI surveys. The problems for the assessment, apart from the lack of understanding about what has driven the apparent outburst in arrowtooth flounder since the early 1980s, are incomplete compositional data and the need to work with an age conversion matrix, and a clear signal from surveys that there is a larger proportion of females than males. Explaining that difference in the model requires assuming differential natural mortality (as done) or allowing the difference to be explained through selectivity and catchability estimation. The choice of using fixed, differential M is appropriate but any consideration of selectivity fits then needs to be careful not to over-interpret resulting ogives too literally as pure selectivity functions. The work done to explore selectivity is appropriate and careful but it is not totally clear if the final, base case assessment model should be that presented in the 2016 SAFE report (Ref 1) or a modified variant using either or both a revised age conversion matrix and/or a change in bounds on selectivity fitting (as in ref 9). From a model fitting perspective, widening the bounds means they are not hit, removing the constraints on parameter estimation, but the resulting fits when the bounds are widened are not wholly convincing for at least two surveys given differences in male and female growth rates. Also, the likelihoods suggest the widening of bounds fits biomass indices somewhat less well. The decision on use of 2016 or a modified assessment might in any case rest on a simple consideration of how the alternative models impact through the estimated fishery selectivity ogives on YPR-related reference points and whether or not they can be reliably estimated to provide management advice.

I **recommend** one simple first calculation. That is, calculate the reference points for the modified model and the 2016 base case (both with the new age conversion matrix). The status determinations will clearly be the same across assessment models, but the YPR-related reference points for each variant will depend on the estimated fishery selectivity and its interaction with the maturity schedule. Given the very different estimated fishery selectivities shown during the review, it is not guaranteed that the YPR-related quantities will be similar. If the management advice is robust, then fine-tuning the assessment further is unnecessary. However, if the advice is not robust then further consideration is needed not just of how to deal with the bounds issue but also whether it is appropriate to use a non-parametric fit for fishery selectivity. Allowing too much freedom in fishery selectivity can result in widely varying fits as adjustments are made. When management advice relies on YPR-related reference points this can create problems not just in providing advice in year x but also consistently through time.

BSAF 4. Recommendations for further improvements to the assessment model.

See recommendations coloured red embedded in text above.

BSAI Flathead Sole (BSAIFS)

A full assessment of BSAI Flathead sole (*Hippoglossoides elassodon*), as the main part a group assessment with Bering flounder (*Hippoglossoides robustus*), was carried out by

McGilliard *et al.* (2016; Ref 6; Chapter 9 of the NPFMC Bering Sea and Aleutian Islands SAFE Report). That assessment and additional exploratory analyses were presented by McGilliard *et al.* (Ref 14). The assessment was accepted by the SSC, which also supported suggestions by the authors for further exploratory work on alternatives to the use of length-based selectivity estimation for the surveys, and the move to implementation using Stock Synthesis to aid in exploratory work.

BSAIFS 1. Evaluation of the ability of the stock assessment model for flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Bering Sea and Aleutian Islands

See under ToR BSAIFS 3.

BSAIFS 2. Evaluation of the strengths and weaknesses in the stock assessment model for Bering Sea/Aleutian Islands (BSAI) flathead sole

See under ToR BSAIFS 3.

BSAIFS 3. Evaluation of alternatives to the current length-based survey selectivity curves used in the assessment

Two streams of information presented suggest survey size-selectivity (one survey, combined sex) is poorly estimated. First, the residual patterns (e.g., Ref 18, Fig 9-17) show strong, systematic size and cohort structure. The residual patterns are strong for both males and females, though perhaps are more pronounced for males – perhaps due to an unaccounted sex difference in natural mortality? Both show substantial positive residuals for the smallest fish (down to 6 cms), suggesting too shallow a selectivity ogive (and an annoying, positive intercept). However, the patterns are largely the same for males and females, suggesting separation of selectivity by sex may not be helpful. Similarly, though the residuals by size are strong, the residuals by cohort suggest that moving to age-based selectivity may also not be a panacea. This is especially the case as the survey age composition data seem to be well fitted in the current model.

Secondly, the retrospective pattern is a concern as outlined in the SAFE report (Ref 6). While the fit to survey biomass is good, there is a strong tendency in the 2016 assessment for observed indices over the past decade to be higher than the model estimates. As data are retrospectively dropped, it is therefore not surprising that the retrospective fits estimate progressively higher final year biomass (and consequentially higher F's to scale to the catches given constant catchability). The retrospective pattern seen in survey selectivity slope and L50 estimates (Ref 14, slide 32), however, is possibly of most concern with both showing strong trends. The L50 trend is striking.

The trend in selectivity estimates suggests that the most fruitful avenue for consideration is likely to be in exploring time varying selectivity (continuous, running [as in EBS pollock, by age], or in blocks). However, all of these require a large number of additional parameters and the data may be insufficient. Also, even if improved fits are achieved (at the expense of parametrisation), evaluating what is best will be problematic, requiring considerable justification in terms of underlying changes in the survey, fish distribution (availability), management history, or perhaps environment-related vulnerability. Justification in purely statistical terms will be technically challenging given each model would be differently parameterised and weighted. Justification in terms of pattern is always subjective and the underlying complexities can confuse.

In order of model tractability, alternative functional form exploration would be simplest, followed by sex-specific selectivity, age-based selectivity, and then time-varying selectivity. Apart from exploring alternative (not logistic) functional forms, this order is the reverse of what might be appropriate as a way of improving the model fit.

It is not obvious that a simple change of functional form would result in "better" selectivity fits, either based on size or on age – it is the very small fish that are over-fit and they would translate cleanly in to a single age group that would also likely be over-fitted. A change in form, e.g. to a dome, would also not of itself remove the retrospective pattern unless it were to filter annual survey size/age compositions in such a way as to remove the apparent recent-decade "bias" in biomass estimation compared to the survey indices. The survey size residuals show some cohort structure, so this is likely feasible in fitting terms even if results would be hard to interpret.

However, while it is perplexing from a modelling perspective that survey selectivity is so poorly fit, the estimates of fishery selectivity seem exceptionally robust with no obvious sensitivities and excellent residuals. Given the robustly estimated fishery selectivity L50 of about 37 cm and steep selectivity slope, plus the maturity ogive and growth relationship for females suggesting most fish are mature before recruitment to the fishery, YPR-related reference points should be robust. Adoption of the 2016 base case model or variants with different survey selectivity fits would not likely result in different stock status determination or management advice.

What potentially matters in terms of advice is the retrospective pattern in biomass, if it feeds through to trending advice on catches. Given, however, the very low actual F relative to F40% and F35%, this is moot and beyond the scope of review.

I am not sure what, if anything, to recommend given the foregoing. It is quite possible to explore functional forms, age- *vs* size-based selectivity, selectivity split by sex, time-varying selectivity, etc. My strong suspicion in this case is that only time-varying selectivity might fit the data better and I would **recommend** this as a starting point for such work. I am doubtful, however, that better fits using time-varying selectivity would necessarily be regarded as fully justified. Given that fishing mortality is very low and stock status very high, and advice emanating from the model may in any case be robust (and in that sense, reliable), it is not clear to me that fuller explorations on survey selectivity are necessary.

BSAIFS 4. Potential evaluation of an equivalent BSAI flathead sole assessment model in Stock Synthesis

This ToR might be interpreted as s*hould an SS equivalent BSAI flathead sole assessment model be developed?* or, *is the SS-implemented BSAI flathead sole assessment model as presented for review equivalent to the existing ADMB-implemented assessment described in Ref 6?*

The answer to the first question depends on whether priority is given to exploring alternative selectivity fitting options. If priority is given, as suggested by the SSC comments in the SAFE report (Ref 6), then the answer is yes. Indeed, from the presentation, it appears this has already been done.

During the review, reasons for developing an SS equivalent model were given (Ref 14):

- *Way to ground-truth dynamics of the 2016 model*

- *Very flexible; many, many options already built in:*
    - *Multiple fleets*
    - *Alternative ways to estimate selectivity*
    - *Time-varying effects available for most parameters*
    - *Stock-recruitment options*
    - *Can estimate growth within the assessment model (better accounts for uncertainty)*
    - *Ageing error definitions*
    - *Etc.*
- *Old Model code is quirky – needs a lot of work to move forward with it.*

I will make no comment on the last bullet point. On the first, I suggest this might better be phrased to say that implementing the model in SS is a way to verify that the assessment is coded as described/intended; I do not see it as a *way to ground-truth* (i.e., validate?) the model. What is true is that SS provides a flexible framework as described, with many of the options immediately available to explore selectivity options as outlined in Ref 6 and at ToR BSAIFS 3, time-varying parameters, etc. As SS can also handle uncertainty estimation/propagation and includes a useful *management layer*, it is a reasonable way to proceed.

On the second interpretation of the ToR, the presentation by McGilliard *et al.* (Ref 14) includes slides on matching a new SS-implemented model with the 2016 assessment. These were discussed during the review. The presentation also includes some slides on exploratory models using SS. These were not examined during the review, but I am relieved to note that the explorations are broadly consistent with my expectations noted at ToR BSAIFS 3 (and acknowledge I should have reviewed them in advance of writing!) I note in particular, however, that while time-varying selectivity (annual or pre- and post-Amendment 80) appear to fit the data better, neither of them, nor dome-shaped selectivity, changes the fit to biomass indices. I would expect all fits still to display systematic retrospective patterns.

On the matching exercise itself, to create an equivalent base model, I see no major difficulties. SS effectively fits survey selectivity by age and a re-scaling on catchability is required to make the results almost equivalent. Diagnostics were not available fully to compare the ADMB and SS-implemented model fits to survey lengths and ages, but visual inspection of the SS outputs suggests age is better fitted than size and that there remains considerable inter-annual variability in the residual patterns (suggesting the potential for time-varying selectivity to be important in improving fits). While the survey selectivity may be visually better fit (on age) than on size (as previously) there are no detailed diagnostics for 2016 and SS model comparisons. Importantly, in matching the models, the biomass fits remain the same and the expectation of strong retrospective behaviour remains.

Overall, the assessment authors have successfully implemented a flathead sole 2016-equivalent assessment in SS. The equivalent assessment has the same features and outcomes as the 2016 assessment and SS can provide a useful framework for exploring alternative selectivity models. The work thus far has been limited but shows results broadly in line with expectations based on examination of fits to the 2016 assessment (above).

Going forward, while time-varying selectivity is apparent, it is not obvious that its inclusion provides any assessment benefits. Judging on the fits to biomass indices, allowing even annually varying selectivity still cannot deal to the likely retrospective problem. Also, using time-varying selectivity for surveys but not the fishery has no impact on projections/advice. Time-varying selectivity would in general be expected for the fishery rather than survey. It is not at all clear why the survey selectivity should be blocked with respect to Amendment 80 and fitting annual selectivities suggests little faith in the survey as an index of abundance.

Ideas for the future presented in the review include i) separation of the pelagic and non-pelagic fisheries; ii) reconsidering the relationship between catchability and temperature; iii) fixing or setting a prior on M and potentially estimating q; iv) including growth estimation within the model; and v) estimating stock-recruitment.

I doubt that (i) will be of importance and would expect selectivity estimation for the non-pelagic fishery to be highly variable. Perhaps, however, separation by fishery could also be accompanied by selectivity blocking of the flatfish/Amendment 80 fishery components? The current model does fit fishery selectivity surprisingly well and separating the fisheries could lead to better characterisation of uncertainty. Item (ii) is interesting. The fitted additive effect was small; a multiplicative effect may be helpful in re-scaling biomass, especially since about 2000 (based on visual inspection). There is potential here to modify the biomass estimates over the past decade and so reduce the retrospective pattern. I note that while the retrospective pattern is clear, it is not especially strong in terms of Mohn's Rho statistic and may not warrant special attention. Also, introducing an annual temperature dependency would potentially confound with annual selectivity fitting. Item (iii) is interesting, with only fixed M considered thus far. Introducing a prior on M or q could be interesting and would certainly help to better characterise uncertainty. Estimating growth within the model (iv) would allow fuller account to be taken of uncertainty but it seems a step too far in a model already struggling to fit size and age-related parameters; I would not afford it high priority. With regard to estimation of stock-recruitment (v) within the model, there is likely more information in the data potentially to estimate M (as at iii) than stock-recruitment steepness; as the two are confounded I think focus on M, if either, would be preferable.

Appendix 1

Materials referenced in SOW

Items 1, 4, 5, and 6 downloaded via provided links. Item 3 sent by e-mail on request. Item 2 unavailable.

1. Spies, I., Wilderbuer, T.K., Nichol, D.G. and Hoff, J,. Palsson, W., 2016. Arrowtooth flounder. *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions*, pp.921-1012. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIatf.pdf
2. Doyle, M., Debenham, C., Barbeaux, S., Buckley, T., Spies, I., Pritle, J., Shotwell, K., Wilston, M., Cooper, D., Stockhausen, W., and Duffy-Anderson, J. In Prep. A full life history synthesis of Arrowtooth Flounder ecology in the Gulf of Alaska.
3. Wilderbuer, T. and Turnock, B. 2009. Sex-Specific Natural Mortality of Arrowtooth Flounder in Alaska: Implications of a Skewed Sex Ratio on Exploitation and Management, North American Journal of Fisheries Management, 29:2, 306-322, DOI: 10.1577/M07-152.1.
4. Wilderbuer, T., J. Ianelli, D. Nichol, and R. Lauth. 2016. Assessment of the Kamchatka flounder stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIkamchatka.pdf
5. NPFMC. 2017. BSAI Introduction. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries
Management Council, Anchorage, AK. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIintro.pdf
6. McGilliard, C.R., Nichol, D. and Palsson, W. 2016. 9. Assessment of the Flathead Sole-Bering flounder Stock in the Bering Sea/Aleutian Islands Regions. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. pp. 1229-1318. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIflathead.pdf

Presentations provided on site

Presentations were provided as follows (items 7-15) plus one on the North Pacific observer program by M. Conception (with no accompanying slides/file).

7. Wilderbuer. Overview of Bering Sea flatfish fisheries and management. (ppt file, 20 slides).
8. Anderl and Matta. Flatfish Age Determination at the Alaska Fisheries Science Center (ppt file, 24 slides).
9. Spies et al. The Bering Sea and Aleutian Islands, and Gulf of Alaska arrowtooth flounder stock assessment (ppt file, 148 slides).
10. Hoff. Eastern Bering Sea Upper Continental Slope Groundfish Bottom Trawl Survey

(ppt file, 33 slides).
11. Wilderbuer, Ianelli, Nichol and Lauth. Assessment of the Kamchatka Flounder stock in the Bering Sea and Aleutian Islands. (ppt file, 33 slides).
12. Laman et al. Aleutian Islands Bottom Trawl Survey (1980 – present). (ppt files, 30 slides).
13. Lauth et al. Eastern Bering Sea Upper Continental Slope Groundfish Bottom Trawl Survey (1980 – present). (ppt file, 29 slides).
14. McGilliard, Nichol and Palsson. BSAI Flathead Sole Complex (ppt file, 74 slides).
15. Haynie. The Amendment 80 fishery: A long-term view on management changes that have impacted this North Pacific multi-species fishery (pdf file, 28 slides).

Other Reference materials

16. Brodziak, J., J. Ianelli, K. Lorenzen, and R.D. Methot (2009) Estimating Natural Mortality in Stock Assessment Applications. August 11–13, 2009 Alaska Fisheries Science Center, Seattle, WA.
17. Lorenzen, K. 1996. Te relationship between body weight and natural mortality in juvenile and adult fsh: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49:627-647.
18. Wilderbuer, T. K., and T. M. Sample (2002) Arrowtooth flounder. In Stock Assessment and Fishery Evaluation Document for Groundfish Resources in the Bering Sea/Aleutian Islands Region as Projected for 2003, p.283-320. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage Alaska 99510.
19. Stark, J. 2011. Female maturity, reproductive potential, relative distribution, and growth compared between arrowtooth flounder (Atheresthes stomias) and Kamchatka flounder (A. evermanni) indicating concerns for management. Journal of Applied Ichthyology. 28(2) 226-230. doi:10.1111/j.1439-0426.2011.01885.x.

Appendix 2

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Fisheries Stock Assessments for Arrowtooth Flounder, Flathead Sole and Kamchatka Flounder

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available. NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Alaska Fisheries Science Center's (AFSC) Resource Ecology and Fisheries Management Division (REFM) requests an independent review of the integrated stock assessments that have been developed for three Bering Sea flatfish species; arrowtooth flounder, flathead sole and Kamchatka flounder. The fishery for these species is managed by the North Pacific Fisheries Management Council. The sum of the Allowable Biological Catches (ABCs) for these three species is 142,529 t in 2017, with catch levels annually set lower than the ABC due to a 2.0 million t harvest cap for all species and constraints due to Pacific halibut bycatch limits and markets. The catch limits are established using Automatic Differentiation (AD) Model software that uses survey abundance data and survey and fishery age and length composition data with a harvest control rule to model the status and productivity of these stocks and set quotas. Having these assessments vetted by an independent expert review panel is a valuable part of the AFSC's review process. The Terms of Reference (TORs) of the peer review and the tentative agenda of the meeting are below.

Requirements for CIE Reviewers

NMFS requires three CIE reviewers to conduct an impartial and independent peer review in accordance with the SOW, OMB Guidelines, and the TORs below. The reviewers shall have working knowledge and recent experience in the application of fisheries stock assessment processes and results, including population dynamics, separable age-structured models, harvest strategies, survey methodology, and the AD Model Builder programming language. Experience with the Stock Synthesis Assessment Model would also be helpful. They should also have experience conducting stock assessments for fisheries management.

Statement of Tasks

- Review the following background materials and reports prior to the review meeting:

Spies, I., Wilderbuer, T.K., Nichol, D.G. and Hoff, J,. Palsson, W., 2016. Arrowtooth flounder. *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions*, pp.921-1012. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIatf.pdf

Doyle, M., Debenham, C., Barbeaux, S., Buckley, T., Spies, I., Pritle, J., Shotwell, K., Wilston, M., Cooper, D., Stockhausen, W., and Duffy-Anderson, J. In Prep. A full life history synthesis of Arrowtooth Flounder ecology in the Gulf of Alaska.

Wilderbuer, T. and Turnock, B. 2009. Sex-Specific Natural Mortality of Arrowtooth Flounder in Alaska: Implications of a Skewed Sex Ratio on Exploitation and Management, North American Journal of Fisheries Management, 29:2, 306-322, DOI: 10.1577/M07-152.1.

Wilderbuer, T., J. Ianelli, D. Nichol, and R. Lauth. 2016. Assessment of the Kamchatka flounder stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIkamchatka.pdf

NPFMC. 2017. BSAI Introduction. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIintro.pdf

McGilliard, C.R., Nichol, D. and Palsson, W. 2016. 9. Assessment of the Flathead Sole-Bering flounder Stock in the Bering Sea/Aleutian Islands Regions. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. pp. 1229-1318. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIflathead.pdf

- Attend and participate in the panel review meeting
    - The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers
- After the review meeting, reviewers shall conduct an independent peer review in accordance with the requirements specified in this SOW, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus
- Each reviewer may assist the Chair of the meeting with contributions to the summary report, if required by the TORs
- Deliver their reports to the Government according to the specified milestone dates

Foreign National Security Clearance
When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The place of performance shall be at the contractor's facilities, and at the Alaska Fisheries Science Center, Seattle, Washington.

Period of Performance
The period of performance shall be from the time of award through June 12, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
No later than April 4, 2017	Contractor provides the pre-review documents to the reviewers
April 18-20, 2017	Panel review meeting

May 8, 2017	Contractor receives draft reports
May 30, 2017	Contractor submits final reports to the Government

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each TOR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $10,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Tom Wilderbuer
Tom.Wilderbuer@noaa.gov
National Marine Fisheries Service,
7600 Sand Point Way, NE, Bldg. 4,
Seattle, WA 98115-6349
Phone: (206) 526-4224

Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Terms of Reference for the Peer Review

Bering Sea and Aleutian Islands Arrowtooth flounder

1. Evaluation of the ability of the stock assessment model for arrowtooth flounder, combined with the available data, to provide parameter estimates to assess the current status of arrowtooth flounder in the Bering Sea and Aleutian Islands.
2. Evaluation of the strengths and weaknesses in the stock assessment model for arrowtooth flounder.
3. Evaluation of the assumption that male natural mortality is higher than female in arrowtooth flounder.
4. Recommendations for further improvements to the assessment model.

Bering Sea and Aleutian Islands Kamchatka flounder

1. Evaluate stock assessment approach to model the Kamchatka flounder resource using three spatially distinct trawl surveys to provide reliable estimates of productivity, stock status, and statistical uncertainty for management advice.
2. Evaluate likelihood profile approach to estimate natural mortality rate (and suggest/provide alternatives?)
3. Evaluate how survey catchability estimates are derived based on assumptions about relative stock distributions.

Bering Sea and Aleutian Islands flathead sole

1. Evaluation of the ability of the stock assessment model for flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Bering Sea and Aleutian Islands
2. Evaluation of the strengths and weaknesses in the stock assessment model for Bering Sea/Aleutian Islands (BSAI) flathead sole
3. Evaluation of alternatives to the current length-based survey selectivity curves used in the assessment
4. Potential evaluation of an equivalent BSAI flathead sole assessment model in Stock Synthesis

Tentative Agenda

TBD

Alaska Fisheries Science Center

7600 Sand Point Way NE

Seattle, WA 98115

April 18-20, 2017 9AM - 5PM

Point of contact: Tom Wilderbuer (tom.wilderbuer@noaa.gov)

Appendix 3

List of non-CIE Participants provided by AFSC

Anne Hollowed	AFSC Status of stocks
Carey McGilliard	AFSC Status of stocks
Ingrid Spies	AFSC Status of stocks
Meaghan Bryan	AFSC Status of stocks
Tom Wilderbuer	AFSC Status of stocks
Sandra Lowe	AFSC Status of stocks
Jim Ianelli	AFSC Status of stocks
Alan Haynie	AFSC Economics program
Jerry Hoff	AFSC Bering Sea groundfish survey
Bob Lauth	AFSC Bering Sea groundfish survey
Dan Nichol	AFSC Bering Sea survey program
Ned Laman	AFSC Aleutian Islands groundfish survey
Beth Matta	AFSC Age and growth program
Delsa Anderl	AFSC Age and growth
Marlon Concepcion	AFSC Observer program
Todd Loomis	Industry (Ocean Peace)

Final Agenda (sent 15th April 2017)

Tuesday April 18th		
9:00	Welcome and Introductions, adopt agenda	
9:15	Overview (species, biology, surveys, fishery, catch levels, ABCs, TACs, bycatch)	Tom
9:30	Bering Sea trawl shelf survey	Bob Lauth
10:00 10:30	Aleutian Islands trawl survey Bering Sea slope trawl survey	Ned Laman Jerry Hoff
11:00	Coffee break	
11:15	Observer Program	Lisa Thompson
11:40	Age Determination	Delsa Andryl
12:00	Lunch	
1:00	Effect of multiple management actions on flatfish fisheries	Alan Haynie
1:30	Bering Sea Arrowtooth flounder	Ingrid Spies
Wednesday April 19th		
9:00	BSAI flathead sole	Carey McGilliard
11:00	Coffee break	
11:20	BSAI flathead sole (continued)	Carey McGilliard
12:30	Lunch	
1:30	Bering Sea Kamchatka flounder	Tom Wilderbuer and Jim Ianelli
Thursday April 20th		
9:00	Bering Sea flatfish discussion	
11:00	Coffee break	
11:20	CIE panel discussion (assessment authors will be available)	
12:30	Lunch	
1:30	CIE panel discussion (assessment authors will be available)	